

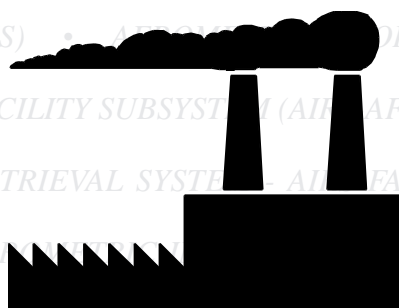


Center for  
Environmental  
Information and  
Statistics

US Environmental  
Protection Agency

# Major Findings from the CEIS Review of EPA'S

## AEROMETRIC INFORMATION RETRIEVAL SYSTEM (AIRS) - AIRS FACILITY SUBSYSTEM (AIRS-AFS) DATABASE



April 2, 1999

---

# Major Findings from the CEIS Review of EPA's AIRS-AFS Database

---

EXECUTIVE SUMMARY .....	1
1. INTRODUCTION .....	2
2. SUMMARY ANSWERS TO REVIEW QUESTIONS .....	3
3. DETAILED ANSWERS TO REVIEW QUESTIONS .....	5
3.1. What does the database cover? .....	5
3.2. Can the database be used for spatial analysis? .....	7
3.3. Can the database be used for temporal analysis? .....	7
3.4. How consistent are the variables over space and time? .....	7
3.5. Can data from AIRS-AFS be linked with information from other databases? .....	8
3.6. How accurate are the data in AIRS-AFS? .....	8
3.7. What are the limitations of AIRS-AFS? .....	9
3.8. How can I get information on AIRS-AFS? .....	9
3.9. Is there documentation on AIRS-AFS? .....	10

## EXECUTIVE SUMMARY

The **Aerometric Information Retrieval System (AIRS)/AIRS Facility Subsystem (AIRS-AFS)** was developed as a national database system for the collection and dissemination of compliance information for all major stationary sources of air pollution in the U.S. AIRS-AFS is operated as a collaborative program between EPA, State and local government agencies that collect and input data in keeping with Federal reporting requirements. The system is co-managed by the Office of Air Quality Planning and Standards (OAQPS) and the Office of Enforcement and Compliance Assurance (OECA) as it contains both emissions and compliance data. OAQPS is primarily responsible for the operation and maintenance of the systems, while the compliance information contained within the systems is used by OECA and the States to carry out their respective programs for improving and maintaining air quality.

This database maintains compliance information on facilities regulated under the Clean Air Act, as well as Titles I, III, IV, and V of the Amendments. The compliance information within the database pertain to major sources of hazardous and criteria air pollutants. A major source is defined as any stationary source or group of stationary sources located within a contiguous area and under common control that emits or has the potential to emit air pollutants above thresholds specified in the Clean Air Act and its Amendments. Nationally required information contained within the database includes facility identification, significant violation status, air programs that regulate the source, pollutants covered by the air programs, and the required actions (inspections, notice of violation, penalties, etc.) within the air programs.

AIRS-AFS contains approximately 100,000 facility records reported by States, 40,000 of which are Federally reportable facilities, with the balance being those tracked to meet State requirements. AIRS-AFS data are collected and updated by State and/or local agencies, and data reporters have direct access to the system at EPA's National Computing Center. The submission process includes checks during data entry to minimize errors. Regions provide quality assurance for the data provided by the States within that Region.

State compliance data coverage within AIRS-AFS is uneven as a result of varying State definitions and requirements, as well as differences in the structure and operation of State data collection, translation, and reporting activities. Analyses using only the nationally required data are more meaningful in situations where consistency is required.

AIRS-AFS data are appropriate for both spatial and temporal analyses. Spatial analyses are possible through utilization of address and ZIP code information which is consistently available. Updates to AIRS-AFS are made quarterly, but some States may be more diligent than others with regard to the timeliness of their updates, which may result in temporal inconsistencies. AIRS-AFS data can be linked to that contained within other databases using the Facility Identification Initiative on Envirofacts. Greater linkage capabilities are forthcoming through AIRS staff involvement with the Sector Facility Indexing Project.

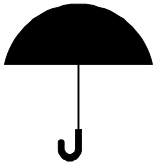
# 1. INTRODUCTION

The EPA developed the **Aerometric Information Retrieval System (AIRS)/AIRS Facility Subsystem (AIRS-AFS)** as a national database system for the collection and dissemination of compliance information concerning all major stationary sources of air pollution in the United States. A major source means any stationary source or group of stationary sources located within a contiguous area and under common control that emits or has the potential to emit air pollutants above thresholds specified in the Clean Air Act and its Amendments. The AIRS-AFS is operated as a collaborative program between EPA, and the State and local agencies that collect and input data in keeping with their Federal data reporting requirements under the Clean Air Act and its Amendments.

Compliance information on facilities affected by the Clean Air Act, including Titles I, III, IV, and V of the Amendments are in the database. Title I refers to the National Ambient Air Quality Standards (NAAQS) set for criteria pollutants. Section 110 of the Clean Air Act of 1970 requires development of State Implementation Plans (SIPS) to provide for the implementation, maintenance and enforcement of the NAAQS for each air quality region in each State. The 1977 Amendments required the implementation of a program for the review of new source locations prior to their construction or modification. SIPS sources as well as New Source Performance Standards (NSPS) sources are found in AIRS-AFS. Title III of the Amendments refers to imposing controls on emissions of hazardous air pollutants (HAPs). Title IV caps SO<sub>2</sub> emissions and lowers NO<sub>x</sub> emissions primarily from major utility emission sources. Title V establishes a comprehensive emissions permit program for major sources, including sources subject to Title III. Specific permit requirements and provisions, as well as State implementation of the program, are set out in Title V. Most of these regulations are based on the use of appropriate control technology. Compliance status and violations are based on the performance of control technology.

AFS became operational in April, 1990, replacing the National Emissions Data System (NEDS), the Hazardous and Trace Emissions System (HATREMS), and the Emissions Inventory System (EIS) for emission point source data reporting. AIRS-AFS also replaced the Compliance Data System (CDS), previously used by EPA and the States for compliance and emission point source data reporting, and the Continuous Emissions Monitoring Subset (CEMS), which was used for the maintenance of point continuous emission monitoring data.

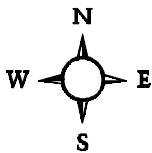
## 2. SUMMARY ANSWERS TO REVIEW QUESTIONS



### 2.1. What does the database cover?

The AIRS-AFS contains compliance data for major sources of air pollutants, including criteria pollutants and hazardous air pollutants. These sources are collectively referred to as Federally reportable facilities. The database includes nationally required information on facility identification; significant violation status; air programs that regulate the source; pollutants covered by the air programs; and the required actions (inspections, notices of violation, penalties etc.) within the air programs. There are compliance-related data for approximately 40,000 of these facilities.

States may use AIRS-AFS or their own systems to maintain data to manage their air programs. States that use AIRS-AFS as their primary system may input more information than the nationally required data. States that maintain their own systems may extract nationally required data from their systems and transfer just that information into AIRS-AFS. Hence there is a variation in the types and levels of information provided by individual States.



### 2.2. Can the database be used for spatial analysis?

Yes. AIRS-AFS records contain geographical identifiers such as Region, State, County, City and ZIP code. Spatial analysis regarding compliance by major sources of air pollutants may be done using any of these identifiers.



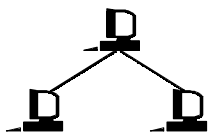
### 2.3. Can the database be used for temporal analysis?

Yes. AIRS-AFS is updated by the States on a quarterly basis if there is new information for that quarter. The records contain information on dates for compliance events such as inspections, notices of violation, administrative orders, and civil referrals. This allows temporal analysis using AIRS-AFS data.



### 2.4. How consistent are the variables over space and time?

The nationally required data contained in AIRS-AFS are consistent. Spatial inconsistencies exist in the additional data provided by States. The level of detail varies between States as a function of State programmatic needs. Additionally, the definitions of violations may differ among States. In instances where spatial consistency is required, the analyses should be limited to the nationally required data. Some inconsistencies are also present in the code values assigned for classifying sources as major or non-major sources. Such differences are due to redefining major sources as per Title V. Finally, some States may also be more diligent than others about timely updates of the database. This may also result in some temporal inconsistencies.



### 2.5. Can data from AIRS-AFS be linked with information from other databases?

Yes. AIRS-AFS data can be linked to other databases through the Facility Identification Initiative (FII) using facility identifying information or geographical identifiers such as Region, State, County, City and ZIP code.



## 2.6. How accurate are the data in AIRS-AFS?

The accuracy of data in AIRS-AFS is subject to a number of factors and influences, including differing State reporting systems, conversion from State to Federal systems, and diligence on the part of State reporters. Quality assurance reviews of AIRS-AFS data are performed by the Regional Offices. Regions use different methods such as

- obtaining data from States, evaluating the data, and inputting the data for the State;
- conducting a series of data retrievals and sending them to States for evaluation; and
- random checks of data inputted by States.

Thus, the level of accuracy differs due to the diligence of both States and Regional Offices.



## 2.7. What are the limitations of AIRS-AFS?

AIRS-AFS is limited by differences in the structure and operation of State data collection, translation, and reporting activities — some States are more conscientious data reporters than others, resulting in uneven data coverage. Analyses using only the nationally required data are more meaningful in situations where consistency is required.



## 2.8. How can I get information on AIRS-AFS?

The public may access AIRS-AFS information through

- Envirofacts : ([http://www.epa.gov/enviro/html/airs/airs\\_query.html](http://www.epa.gov/enviro/html/airs/airs_query.html)), and
- AIRSweb: (<http://www.epa.gov/airsweb/sources.htm>)

Most of the facility and emissions data in AIRS-AFS are available to any person or organization with access to the EPA National Computer Center.

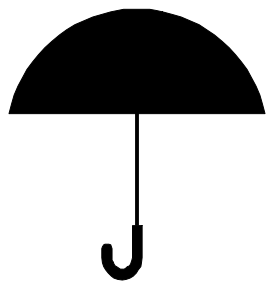


## 2.9. Is there documentation on AIRS-AFS?

There are a number of user documents available from the Office of Air and Radiation and the Office of Air Quality Planning and Standards.

(<http://www.epa.gov/ttn/airs/airsmans.html>)

### 3. DETAILED ANSWERS TO REVIEW QUESTIONS



#### 3.1. What does the database cover?

The AIRS Facility Subsystem (AIRS-AFS) database contains compliance data on a number of pollutants for major stationary sources regulated by the EPA, and State and local agencies. Facilities regulated under the Clean Air Act as well as Titles I, III, IV, and V of the Amendments are included in the database. The database therefore includes sources regulated under the acid precipitation, Title V permits, SIPS, Chlorofluorocarbon (CFC) Tracking, Prevention of Significant Deterioration (PSD), New Source Review (NSR), National Emissions Standards for Hazardous Air Pollutants (NESHAP) and New Source Performance Standards (NSPS) programs. It covers a variety of air pollutants including criteria pollutants and hazardous air pollutants or air toxics. Compliance status and violations are based on the performance of the control technology. AIRS-AFS contains approximately 100,000 facility records reported by States. These include facilities that are Federally reportable and those tracked by States for their own requirements. For the Federally reportable facilities, AIRS-AFS contains compliance-related data for approximately 40,000 facilities. Sources may be classed as major, minor, or synthetic minor, with reporting required for major sources.

EPA defines major sources as

1. Air toxics sources, as defined in Section 112 of the Act (Title III of the Amendments), with the potential to emit 10 tons-per-year (TPY) of any hazardous air pollutant, 25 TPY of any combination of hazardous air pollutants, or a lesser quantity of a given pollutant if the Administrator so specifies;
2. Sources of air pollutants, as defined in Section 302 of the Act, as amended, with the potential to emit 100 TPY of any pollutant; and
3. Sources which are subject to the nonattainment area provisions of Title I, Part D, with the potential to emit pollutants such as ozone precursors including volatile organic compounds (VOC's) and nitrogen oxides ( $\text{NO}_x$ ), carbon monoxide, and particulate matter, in quantities that vary according to the severity of nonattainment.

Major sources are defined according to their potential to emit rather than actual emissions. A facility's potential to emit is the larger of actual emissions or the maximum legally allowed emissions. A synthetic minor source is an air pollution source that has the potential to emit air pollutants in quantities at or above threshold levels but has accepted federally enforceable limitations to keep the emissions below such levels.

The number of specific data variables that may potentially be reported to AIRS-AFS is immense. A data dictionary of several hundred pages accompanies the

AIRS-AFS database. This dictionary describes the full range of variables included in the database structure. However, States are only required to report nationally required data, though many of them report on more. The most frequently updated records within the database include date of last inspection, compliance status, and plant name.

### **Who Must Report?**

States report most of the data maintained in AIRS-AFS. EPA Regional offices report surveillance and enforcement activity undertaken directly by EPA.

### **How are data reported?**

AIRS-AFS data are collected and updated by State and/or local agencies. Data reporters have direct access to the EPA National Computer Center and AIRS. This submission process includes edit checks to reduce the likelihood of erroneous data entry due to keying mistakes. There are two mechanisms through which States can enter data into AIRS-AFS: batch and online.

For the batch update process, States create a file of transactions, transmit the file to the EPA mainframe computer, and submit a batch job. AIRS software then validates the transactions and uses them to update the AIRS-AFS database.

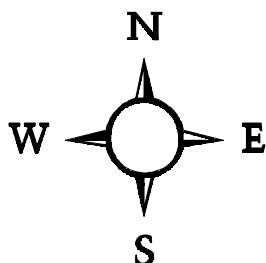
The online update process uses formatted screens to acquire new or modified data. The same validations are performed online as in batch updating, and immediate feedback is provided to diagnose any errors detected in the values entered into the screens. When all the necessary changes or additions have been entered on the screens, the software updates the AIRS-AFS database.

### **Data Elements**

Federally required compliance data cover five basic categories:

1. Identification: Source name, Region, State, County, Source number, Street, City, ZIP code, Standard Industrial Classification (SIC) code to identify the source's primary product, and a Government Ownership indicator to identify government facilities.
2. Significant Violator Status: An indicator showing whether a source is currently a significant violator of any applicable air regulations.
3. Air Programs: Pertinent air programs under which a source is regulated, as well as an indication of the operating status of the source.
4. Pollutants Within Air Programs: Criteria pollutants regulated by the air program and emitted by the source; the classification (major, minor, synthetic minor) for each pollutant; for State Implementation Plan (SIP) sources, an indicator showing whether the source is located within an attainment or nonattainment area, and if in a nonattainment area, the severity of that area; and compliance information.

5. Actions Within Air Programs: Activities which may be recorded at sources include inspections; Notices of Violation; Administrative Orders and any related penalty amounts; and Civil Referrals and any related penalty amounts. Required information includes action number, national action type, and date achieved.



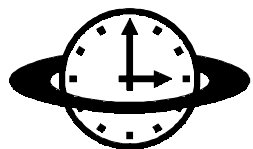
### 3.2. Can the database be used for spatial analysis?

Yes, AIRS-AFS can be used for spatial analysis. Every facility within the database is identified by its street address and ZIP code. In addition, some facilities may be spatially identified by latitude and longitude data if provided by the States. It may be necessary to limit spatial analysis to nationally required data because substantial variation exists in the amount of information reported by States for other data fields.



### 3.3. Can the database be used for temporal analysis?

AIRS-AFS data updates occur on a quarterly basis. Since dates are included for the actions within the air programs, it is possible to conduct temporal analysis. The diligence of the States in reporting the information into AIRS-AFS on a timely basis must also be considered while conducting temporal analysis.



### 3.4. How consistent are the variables over space and time?

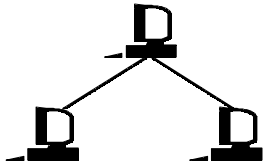
There are several factors to be considered when reviewing the spatial and temporal consistency of AIRS-AFS data. AIRS-AFS is consistent for the nationally required data. States use either AIRS-AFS or their own systems to track the State level programs and to meet Federal requirements. States that use AIRS-AFS as their primary systems tend to input more information than other States. Data other than the nationally required data include information on emissions estimation methods, monitors, control technology, violations of State requirements, etc. Depending on their use of AIRS-AFS for managing State programs and activities, the level of detail varies between States, leading to spatial inconsistencies.

States tend to have their own air programs and regulations beyond those mandated by EPA. The requirements for compliance may differ between States resulting in a difference in the definitions of violations. This inconsistency has to be considered in spatial comparisons using data which are not federally required. In instances where spatial consistency is needed, the analysis should be limited to nationally required data.

There are also some inconsistencies in the code values assigned for classifying sources as major or non-major sources. Currently, there is an on-going process to change classifications as sources are permitted under Title V. As they are permitted, the sources are classified as major or minor per provisions of Title V. However, sources which have not yet been permitted under Title V (a majority of

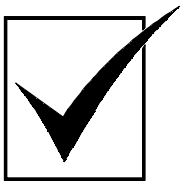
sources to date) generally will still be classified according to older definitions that differ from those in Title V.

States are required to provide quarterly updates when there is reportable compliance information for that quarter. However, some States may also be more diligent than others about timely updates of the database. This may result in some temporal inconsistencies as well.



### **3.5. Can data from AIRS-AFS be linked with information from other databases?**

Yes, AIRS-AFS compliance data can be linked to that contained within other databases in Envirofacts through the Facility Identification Initiative (FII). This improves public access to EPA's environmental data by promoting linkage among EPA's major data sources. FII is a project designed by the EPA to standardize identification information on facilities subject to Federal environmental reporting requirements. It allows users to obtain information from different databases using facility identifiers or geographical identifiers. In addition, AIRS-AFS staff are involved in the Sector Facility Indexing Project (SFIP), which promises to allow even greater ability to link data.



### **3.6. How accurate are the data in AIRS-AFS?**

The accuracy of AIRS-AFS compliance data varies, depending on several factors. Compliance data in AIRS-AFS are collected from several sources, including State computer systems, engineers, inspectors, facility site managers, etc. The data must pass edit checks for entry into the database. Data are checked for plausibility, but not necessarily absolute accuracy. AIRS-AFS is a table-driven system, ensuring that valid values are recorded. A universal interface is being developed to make reporting easier and to eliminate some inconsistencies across States.

The charter of AIRS-AFS is to track all required Federally reportable compliance data. The SFIP, a large-scale review of Agency compliance databases, found that data accuracy is over 95% for federally required compliance data in AIRS-AFS. The States may use AIRS-AFS to track State required compliance data, and local agencies are also free to use AIRS-AFS for tracking compliance data. Data completeness varies from State to State, depending on methods employed to report federally required data. Several reports are available to help system users assess data completeness, e.g., the AIRS-AFS Missing Information Report and the AIRS-AFS Data Completeness Reports. Additional checks for logical inconsistencies are slated to be incorporated, thus strengthening the logical consistency of the database.

Quality assurance review of AIRS-AFS compliance data is performed by the Regional Offices. Some Regions obtain the data from States, then evaluate and input them into the AIRS-AFS system from the Regional Office. Some other Regions have a systematic procedure whereby they retrieve data from the

system and require evaluation and validation by the States. Regions may also randomly check the data provided by the States within that Region.



### **3.7. What are the limitations of AIRS-AFS?**

AIRS-AFS's primary limitation stems from reliance on the States to provide compliance data. Despite the existence of guidance, workshops and conferences on AFS data systems, there are continuing inconsistencies. Therefore, one State's information may not match another's, leading to variation in the database. AIRS-AFS is limited by the conscientiousness of State data collection and reporting efforts. Air enforcement audits conducted by the Office of Inspector General have found problems in State identification and reporting of significant violators of the Clean Air Act (Report # E1GAE7-03-0045-8100244; [www.epa.gov/oigearth/audit/list998/8100244e.htm](http://www.epa.gov/oigearth/audit/list998/8100244e.htm)). It is apparent that some States do a better job of collecting and reporting their data than others, and this disparity leads to 'holes' in the database, which limits the ability to conduct analyses. There are continuing efforts to improve the quality and consistency of the compliance data.



### **3.8. How can I get information on AIRS-AFS?**

#### ***General Information***

General information on AIRS-AFS may be obtained from:  
AIRS/AFS Information Request  
EPA, Office of Air Quality Planning and Standards  
Information Transfer and Program Integration Division (MD-12)  
Research Triangle Park, NC 27711

AIRS/AFS Helpline: (800) 367-1044

#### ***Electronic Media***

AIRS-AFS information is available through the Envirofacts electronic warehouse at

[http://www.epa.gov/enviro/html/airs/airs\\_query.html](http://www.epa.gov/enviro/html/airs/airs_query.html)

and through AIRSWeb at

<http://www.epa.gov/airsweb/sources.htm>

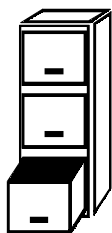
Most of the actual data in AIRS-AFS are available to any person or organization with access to the EPA National Computer Center. There are three ways to retrieve facility data from the AIRS-AFS database: Online Browse; Fixed Format Reports; and Ad hoc reporting.

The AIRS-AFS Online Browse option displays data online. Users specify criteria for data selection (such as plant name, plant identification number, and

desired year(s) of emissions) and AIRS software retrieves and displays the data in formatted screens.

AIRS-AFS Fixed Format reporting generates printed reports and/or data files. Users specify criteria for data selection and sorting, and choose options that affect report format or content. The AIRS software automatically submits a batch job to produce the requested report and/or data file. Many report formats are available, and selection criteria are flexible enough to meet most information needs.

If the Standard Batch Retrieval and Online Browse facilities do not satisfy a user's requirements, experienced users can define supplemental reports via the AIRS-AFS Ad Hoc Batch Retrieval option. Users specify data fields to be included in the report and/or output data file. To do this, however, one must have a thorough understanding of the database organization and data field names.



### **3.9. Is there documentation on AIRS-AFS?**

#### ***Statutory Authority***

- PL 101-549 Clean Air Act Amendments of 1990, Title V Permit Regulations
- PL 97-375 Clean Air Act Amendments of 1977
- PL 91-604 Clean Air Act of 1970

#### ***Internal Consistency***

- AIRS-AFS Data Dictionary, version 2.4
- AIRS-AFS Missing Information Report
- AIRS-AFS Data Completeness Reports

A number of documents are also available online at

<http://www.epa.gov/ttn/airs/airsmans.html>